Learning Triadic Influence in Large Social Networks

Chenhui Zhang*, Sida Gao*, Jie Tang*, Tracy Xiao Liu†, Zhanpeng Fang*, Xu Cheng‡
*Department of Computer Science and Technology, Tsinghua University
†Department of Economics, School of Economics and Management, Tsinghua University
‡Tencent Corporation
{ch-zhang15, gsd13}@mails.tsinghua.edu.cn, jietang@tsinghua.edu.cn, liuxiao@sem.tsinghua.edu.cn
fzp1990@gmail.com, alexcheng@tencent.com

***Abstract*—Social influence has been a widely accepted phenomenon in social networks for decades. In this paper, we study influence from the perspective of structure, and focus on the simplest group structure—triad. We analyze two different genres of behavior: *Retweeting* on Weibo[1] and *Paying* on CrossFire (CF)[2]. We have several intriguing observations from these two networks. First, different internal structures of one's friends exhibit significant heterogeneity in influence patterns. Second, the strength of social relationship plays an important role in influencing one's behavior, and more interestingly, it is not necessarily positively correlated with the strength of social influence. We incorporate the triadic influence patterns into a predictive model to predict user's behavior. Experiment results show that our method can significantly improved the prediction accuracy.**

***Index Terms*—Triadic influence; Social influence; User modeling; Social network**

I. Introduction

Social influence occurs when one's opinions, emotions, or behaviors are affected by others [1]. In this paper, we study social influence from structure level, with a particular focus on the triadic structure. The problem is called *Triadic Influence Analysis*. The reason that we focus on triads is that triad is the simplest group structure in social networks as well as the cornerstone for studying network formation [2], [3]. Employing two large networks: Weibo and CrossFire, we systematically investigate the problem of triadic influence analysis. Our experimental analysis verifies the existence of different triadic influence patterns in the two social networks. We further incorporate the triadic influence to predict user's online behavior and the performance can be significantly improved up to 5-13% compared to various baseline methods.

II. Data and Observation

We study the triadic influence analysis problem on two real-world networks: Weibo and CrossFire. The **Weibo** network consists of 1,787,443 users and 413,503,530 directed relationships. We extracted user attributes including gender, verification status, #reciprocal, #followers, #followees and #microblogs. The action we consider here is retweeting—i.e., when user A tweets a message, will the followers retweet this message as well? The **CrossFire** (CF) network consists of 1,779,270 users and 20,542,973 undirected relationships. We gathered user attributes which contains cf_friend, cf_master, cf_member, reg_date, login_date and user_level. We consider the buying behavior in CF—i.e. will a free user become a payer when s/he plays with some paying users?

To describe a user's neighborhood structure, one's neighbors are grouped into positive and negative while the relationships are grouped into strong and weak. We define two feature sets, say Neighborhood features and Triadic features. The former consist of the number of strong-tied positive neighbors, weak-tied positive neighbors, strong-tied negative neighbors and weak-tied negative neighbors. Triadic features consist of the number of 30 types of triads which are categorized according to the neighbor's label and the strength of relationships between users. Table II lists all types of triads. The dashed line indicates a weak tie and the solid line indicates a strong tie. The red circle denotes a positive neighbor, while blue circle denotes a negative neighbor.

We conduct OLS analysis on aforementioned features. We first test the Neighborhood feature set, and then rerun the analysis of edge features along with Basic features. We further tested triadic features while controlling for all of the basic and edge features. Table II and Table I summarize the results. (The results of Basic Features are omitted due to page limitation.) As for triadic features, we have several intriguing observations. Most of the open triads result in negative coefficients, suggesting that users who have more diverse followees are less likely to retweet. Strong tie also implies a strong influence in general. The relationship between B and C also plays an important role in influencing the retweeting behavior of A when compared the Triad 0 with Triad 1.

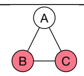
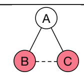
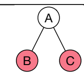
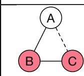
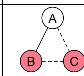
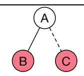
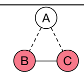
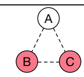
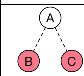
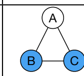
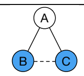
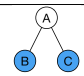
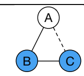
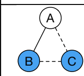
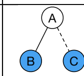
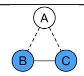
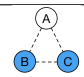
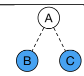
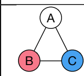
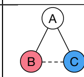
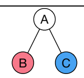
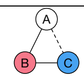
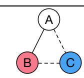
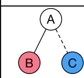
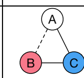
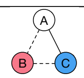
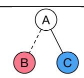
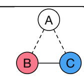
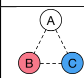
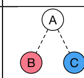
TABLE I
Regression Analysis Neighborhood Features

	Edge	B+E	B+E+Triads
#PosStrong	0.0840*** (0.002)	0.0544*** (0.002)	0.0602*** (0.006)
#PosWeak	0.1268*** (0.003)	0.0700*** (0.002)	0.0679*** (0.012)
#NegStrong	0.0355*** (0.001)	0.1670*** (0.002)	0.0490*** (0.009)
#NegWeak	-0.0490*** (0.001)	-0.0040** (0.002)	0.0201 (0.023)
R^2	0.041	0.242	0.301

[1]http://weibo.com, is the largest microblogging service in China.
[2]CrossFire is an online first-person shooter released in China by Tencent.

IEEE/ACM ASONAM 2016, August 18-21
2016, San Francisco, CA, USA

TABLE II
REGRESSION ANALYSIS FOR 30 KINDS OF TRIADS

No.	Triad	Coef	No.	Triad	Coef	No.	Triad	Coef	No.	Triad	Coef	No.	Triad	Coef
1		0.0827*** (0.003)	2		0.0110*** (0.004)	3		-0.0543*** (0.003)	4		0.0004 (0.004)	5		0.0429*** (0.003)
6		-0.0587*** (0.003)	7		0.0205*** (0.002)	8		0.0313*** (0.002)	9		-0.0283*** (0.003)	10		0.0168*** (0.002)
11		-0.0091*** (0.002)	12		0.0861*** (0.004)	13		-0.0563*** (0.001)	14		-0.0221*** (0.002)	15		0.0157*** (0.003)
16		0.0167*** (0.003)	17		0.0164*** (0.003)	18		-0.0184* (0.010)	19		-0.0263*** (0.002)	20		0.0066*** (0.002)
21		-0.0001 (0.002)	22		0.0099*** (0.002)	23		0.0534*** (0.002)	24		-0.0054** (0.002)	25		0.0089*** (0.001)
26		-0.0783*** (0.002)	27		0.0818*** (0.002)	28		0.0130*** (0.002)	29		0.0494*** (0.002)	30		-0.0772*** (0.003)

III. EXPERIMENT

In this section, we incorporate the triadic structures into user behavior prediction. We adopt Logistic Regression, i.e,. the value of $y_i, i = 1, 2, \ldots, N$ is predicted as:

$$P(y_i = 1|X_i) = \frac{1}{1 + \exp(\alpha X_i + \beta)} \quad (1)$$

where X_i is the feauture vector of the i-th sample. α and β are the parameters and can be learned by maximizing a likelihood objective function defines as:

$$\mathcal{O}(\alpha, \beta) = \prod_{v_i \in V_P} P(y_i = 1|X_i) \prod_{v_i \in V_N} P(y_i = 0|X_i) \quad (2)$$

where V_P and V_N represent the set of positive samples and negative samples respectively. We first train the LRC using the aforementioned basic features separately, i.e., LRC-B, LRC-N and LRC-T, and compare their performance. We further use different combinations of feature sets, which results new comparison methods such LRC-BN, LRC-BNT.

Table III summarize the results. Although Neighborhood features are much less powerful than Basic features when used alone, it can improve the performance significantly (+6.94% (Weibo) and +0.94% (CF) in AUC) when they are combined together. Using Triadic features alone is almost as good as combining Basic and Neighborhood features. Moreover, the Triadic features can further improve the performance by +5.87% on Weibo and +0.52% on CF in AUC when augmented to the LRC-BN, which corroborates that Triadic features can describe the neighborhood network structures in a more refined way.

IV. CONCLUSION

In this paper, we study the social influence from the perspective of triadic structures in one's egocentric network. We classify triadic structures and conduct the OLS analysis to confirm the existence of triadic social influence. Experimental results indicate that the predictive power increases significantly by adding triadic features.

TABLE III
PERFORMANCE OF BEHAVIOR PREDICTION ON TWO DATASETS. (%)

Dataset	Model	Precision	Recall	F1	AUC
Weibo	LRC-B	64.54	71.19	67.70	71.90
	LRC-N	58.46	52.27	55.19	61.53
	LRC-T	67.59	67.25	67.42	74.09
	LRC-BN	69.27	73.83	71.47	77.31
	LRC-BNT	**73.16**	**76.46**	**74.78**	**81.85**
CF	LRC-B	69.63	**72.27**	70.92	77.62
	LRC-N	71.35	57.06	63.41	72.64
	LRC-T	71.52	58.85	64.57	73.56
	LRC-BN	71.37	70.27	70.81	78.35
	LRC-BNT	**71.53**	71.02	**71.27**	**78.76**

REFERENCES

[1] M. Deutsch and H. B. Gerard, "A study of normative and informational social influences upon individual judgment," *Journal of Abnormal and Social Psychology*, vol. 51, pp. 629–636, 1955.

[2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, pp. 824–827, 2002.

[3] H. Huang, J. Tang, L. Liu, J. Luo, and X. Fu, "Triadic closure pattern analysis and prediction in social networks," *IEEE TKDE*, vol. 27, pp. 3374–3389, 2015.